



King's Research Portal

DOI:

[10.1086/691576](https://doi.org/10.1086/691576)

Document Version

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Alvarez, M. (2017). [Book Review] Scott Sehon, Free Will and Action Explanation: A Non-causal, Compatibilist Account. *ETHICS*, 127(4), 958-963. <https://doi.org/10.1086/691576>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

stunningly beautiful—let us imagine you are in Kamouraska, watching the sun set over the St. Lawrence River—we surely should welcome that this person was filled with joy, in spite of what seems a lack of sensitivity. Indeed, it might well be the case that this particular sunset is very significant to that person in part because she does not tend to enjoy this kind of beauty. Because this intuition depends on the availability of objective evaluative features, it is likely to be controversial. But then, it might well be true nonetheless.

ACKNOWLEDGMENT: Thanks to Paul Boswell and to Carolyn Price for their helpful comments.

CHRISTINE TAPPOLET
Université de Montréal

Sehon, Scott. *Free Will and Action Explanation: A Non-causal, Compatibilist Account*. Oxford: Oxford University Press, 2016. Pp. xi+235. \$74.00 (cloth).

This is an ambitious, wide-ranging, well-informed, and carefully argued book. Its main aim is to demonstrate that we have free will and moral responsibility, that is, that we have the responsibility of “desert,” which makes us appropriate subjects for praise or blame and therefore deserving of punishment or reward. A second, radical aim is to reshape those debates, since, Scott Sehon claims, they are “largely informed by a false presupposition: the causal theory of action” (3), and to persuade the reader that his teleological theory is the ideal alternative to the causal account of action explanation.

The book’s structure reflects this dual aim. After chapter 1, which explains the rationale for the book, the remaining twelve chapters fall in two parts. In part 1, comprising chapters 2–7, Sehon outlines and defends the claim that rational explanations of human behavior are irreducibly teleological, and not causal. Part 2, comprising chapters 8–13, applies this teleological account to the problem of free will and responsibility. This yields a view Sehon calls “non-causal compatibilism,” which, he argues, undermines incompatibilist arguments and is preferable to its compatibilist rivals. Moreover, it is claimed in the last chapter that non-causal compatibilism has the resources to explain the puzzle arising from the fact that compatibilism and incompatibilism both “seem to have something going for them” (215).

Sehon begins by arguing against skeptics of all stripes that the question whether we have free will and desert responsibility is genuine and pressing for, he says, without such responsibility, our practices of praise and blame, punishment and responsibility, and the attitudes that he sees as underlying those responses (indignation, gratitude, shame, etc.) cannot be properly justified. Attempts (e.g., by Derk Pereboom) to find alternative justifications fail: in the absence of genuine desert, Sehon claims, those practices and attitudes would be rendered “hollow and manipulative,” and the world would be “deeply depersonalized” (20).

The central idea of the teleological account is that we explain agents’ actions by constructing theories that make sense of them in terms of rationality, in-

interpreting agents so that they come out to be as rational as possible given certain constraints. To do this, Sehon says, we assume that agents (i) “act in ways that are appropriate for achieving their goals” and (ii) “have goals that are of value,” in both cases relative to “their circumstances, epistemic situation, and intentional states” (27). Moreover, in interpreting agents, (iii) we try to make sense of their overall behavior, and not just of actions in isolation, and (iv) we “also take into account what we believe the agent would do in related circumstances” (28). In other words, we explain actions by attributing to agents both goals and beliefs about appropriate ways to achieve them, and in so doing we are constrained by judgments about the value and consistency of agents’ goals and the coherence of their behavior considered diachronically and counterfactually. (It is somewhat ironic that Sehon’s theory is explicitly closely modeled on Davidson’s views on interpretation and his Principle of Charity [Donald Davidson, *Inquiries into Truth and Interpretation* (Oxford: Clarendon, 1984)], given the latter’s role in establishing the causal theory of action explanation as the dominant orthodoxy.)

A second core idea of the theory is that teleological explicability, rationality, freedom, responsibility, and (less plausibly) agency are not “an all-or-nothing affair” but come in degrees. I’ll come back to how this idea plays out in the free will debate. For now, the point helps one to see how Sehon responds to the immediately obvious objection that people are often irrational and act irrationally: since the theory allows for degrees of rationality, it can accommodate occasional irrationality by all, and even frequent and intense irrationality by some. Sehon argues—convincingly, in my view—against claims by social scientists (e.g., Daniel Kahneman, J. L. Knetsch, and R. H. Thaler, “Experimental Tests of the Endowment Effect and the Coase Theorem,” *Journal of Political Economy* 98 [1990]: 1325–48) that we are all in fact systematically irrational.

He also engages carefully with other objections directly targeted at his account. Reasons of space preclude me from commenting on them all, so I shall confine myself to an objection raised by Alfred Mele about agent manipulation. The issue boils down to whether an agent whose limbs are caused to move (e.g., by Martians, or some unusual mechanism) in ways that realize his intentions and accord with his goals and beliefs could rightly be said to be acting. In such cases the teleological account dictates that the agent is acting intentionally, since his ‘behavior’ can be rationalized appropriately. But, Mele urges, the nature of the manipulation suggests that the agent is not acting, never mind acting intentionally.

Sehon’s response is that, insofar as the manipulation is really perfectly responsive to the agent’s psychology, the agent is indeed acting: the “godlike reliable” manipulating device (and associated agents), he suggests, are just a causal mechanism through which the agent executes his intentions, which is, admittedly, more oblique and unusual than, but nonetheless not different in the relevant sense from, the causal mechanisms through which we, ordinary agents, execute ours. Sehon seeks to buttress this response by asking us to consider how we would judge the agent if what he did through such manipulation was morally significant, for instance, shooting someone dead. Surely, he says, “in this scenario, we would not let [the agent] go free, nor would we merely charge him with having had a *plan* to commit murder” (58).

Sehon mentions Malebranche in this context, and appropriately so since this is in effect a modern-day version of the eighteenth-century dispute about

whether, as Thomas Reid put it, the possibility that our volitions have no “physical effect upon the nerves and muscles” and may be, rather, “only an occasion of their being acted upon by some other efficient” (Thomas Reid, *Essays on the Active Powers of the Human Mind* [original title: *Essays on the Active Powers of Man*, 1788; repr., Cambridge, MA: MIT Press, 1969], 50) undermines our status as genuine agents. Reid’s own view was that “the man who knows that such an event depends upon his will, and who deliberately wills to produce it, is, in the strictest moral sense, the cause of the event; and it is justly imputed to him, whatever physical causes may have concurred in its production” (Reid, *Essays*, 51).

So Reid and Sehon agree that what matters for moral assessment is what one intends and decides to do, regardless of the causal mechanisms through which the desired outcomes are brought about. However, emphasizing that, in the imagined scenarios, the agent would be the cause of the event “in the strictest sense,” as Reid puts it, and thus morally blameworthy, does not address the worry that such an agent would not, strictly speaking, be acting. It is only if one accepts Sehon’s other point, that agents do cause the relevant outcomes, albeit in unusual ways, that one will think that Sehon’s response succeeds.

In the last three chapters of part 1, Sehon goes on the offensive, offering a series of trenchant objections to the causal theory. Chapter 5 argues that causal accounts cannot explain why, as they claim, we use rationalizing principles in order to identify causes in reason explanations but “we would never use rationalizing principles to determine the cause of an event” (74) in any other context, since “this would be a laughably silly method” (89) to use in those contexts. The conclusion is that in fact “we do not treat rationalizability as a mere heuristic for finding the right cause” (88)—where by “right cause” Sehon means right physical cause. This is evidence that “reason explanation is not causal but *sui generis*” (89). Chapter 6 is devoted to the familiar problem of deviant causal chains, where Sehon gives convincing arguments that several attempts to solve this problem fail. This failure shows that rationalizations, which presumably everyone agrees are teleological, cannot be analyzed in terms of goals, beliefs (or other mental states), and causation. It does not show that rationalizations are not also causal in character (see below), though the failure is instructive. Chapter 7 turns to a different, original, and subtle objection, which goes roughly as follows. The causal theory implies that commonsense psychology is committed to the view that “mental states, if they exist at all, are identifiable with physical states” (110). But commonsense psychology is not committed to that view. Therefore, the causal theory is false.

Though the arguments in these three chapters are careful and generally persuasive, there is a question about their scope. Sehon’s target is a causal theory that holds that in identifying the cause of an action (the reason), we are thereby identifying the physical cause(s) of the bodily motions involved in the action. And Sehon seems to see this as committing the causal theorist to a particular view about the significance of the theory, namely, that identifying a physical cause is a first step in a project of reducing, or even translating, psychological explanations to physical explanations. No doubt some causal theorists are so committed and motivated. But other causal theorists may see the significance of their position as lying elsewhere. They may agree with the would-be reductionist that rational ex-

planations of human behavior are causal explanations, but insist that they are irreducible and *sui generis*, since they are constrained by principles of rationality, and no other causal explanations are so constrained. That, after all, was Davidson's position. Again, John Hyman (*Action, Knowledge, and Will* [Oxford: Oxford University Press, 2015]) argues that explanations of intentional action are causal but also irreducibly teleological, on the grounds that desires are dispositions that are manifested in goal-directed behavior. I am not claiming that these causalists are right, only that these versions of the causal theory seem to fall outside the target range of views that Sehon's arguments are intended to bring down, since he focuses on views that assume that giving reason explanations is, as he puts it, ultimately a heuristic device for finding the corresponding physical explanations. But that is a project about reducing psychological to physical explanations that a causalist needn't endorse, and which Davidson explicitly disavowed through his commitment to anomalous monism.

I now turn to Sehon's discussion of free will and moral responsibility in part 2, where he deploys his teleological account of action explanation to articulate a compatibilist view of free will and determinism. The discussion covers an impressive range of issues, and it is impossible to do justice to its complexity here. I think that Sehon is most convincing in arguing that when we rationalize actions, we take no account of determinism, whether the actions we explain are more or less rational, ordinary actions, or extraordinary ones, such as the actions of psychopaths or the brainwashed, coerced actions, or those done in Frankfurt-style scenarios. Peter Strawson made this general point ("Freedom and Resentment," *Proceedings of the British Academy* 48 [1962]: 1–25), and Sehon claims that his theory can explain why, since, on the teleological account, "whether a behaviour is free comes down to whether it is *teleologically* explicable" (174), not whether it was caused at all, let alone whether its occurrence could be subsumed under a deterministic causal law. If that is right, then, he also argues, the teleological account undermines certain arguments for incompatibilism that focus on "how our behaviours have been *caused*" (174). This is a suggestive claim, plausible on the face of it, though, of course, its plausibility depends on whether Sehon's teleological account of free action is satisfactory. It is not clear to me that it is.

For Sehon, free actions can be straightforwardly identified with "actions for which we are responsible = intentional actions = goal-directed actions" (129). This, together with the earlier claim that goal-directedness and rationality are a matter of degree, allows Sehon to argue that freedom and responsibility also come in degrees. He summarizes the view as follows: "Behaviors that are less than fully rational are therefore less fully cases of action, and thus less free. Accordingly, an agent's responsibility is at least mitigated somewhat if the behavior was not rational" (145).

Sehon insists that this is preferable to a view that says that freedom and responsibility are a matter of "all or nothing." Perhaps. But the identifications are not unproblematic. For instance, there are things for which we are morally responsible that are not, in any plausible sense, goal directed: negligent but unintentional omissions, and failures to care for the right things or care for them in the right way, for example. Moreover, animals and small infants are capable of goal-directed and, by the teleological account's criteria, rational behavior: they act in ways that are ap-

appropriate for achieving their valued goals, given their circumstances, epistemic situation, and intentional states. Sehon claims this as a virtue of the account because it allows us to see how the degree of rationality increases with the increase of psychological sophistication, either within the development of an individual human or in “a spectrum of different sorts of animals” (139–40; see also 30). But if the actions of animals and infants are goal directed, then it would seem that, according to the teleological theory, they are also free actions for which their agents are morally responsible, even if to a far lesser degree than fully developed adult humans. This is implausible (and, I assume, not a consequence Sehon would welcome): concerning moral responsibility, the difference between us, on the one hand, and animals and infants, on the other, is not a matter of degree but of kind: infants and animals are not proper subjects of moral responsibility.

Finally, I am not convinced that the notion of rational action can do the work that the teleological theory of freedom and moral responsibility requires it to do. Sehon uses both “subjective” and “objective” criteria for rational actions. Thus, he suggests that, relative to his beliefs and desires, the actions of a bank robber (Robert Alton Harris, a real-life example discussed in Gary Watson, “Responsibility and the Limits of Evil,” in *Responsibility, Character, and the Emotions* [Cambridge: Cambridge University Press, 1987], 256–86) who brutally kills two innocent young witnesses and then eats their unfinished burgers—he was hungry!—are rational. However, the objective criterion “makes it more difficult to rationalize some of his behaviors” since he has “subhuman values: he is callously indifferent to the value of human life” (166). Many (e.g., Hume) have been skeptical that being rational is at all a matter of having objectively good values. But putting that aside, the theory has some unwelcome consequences. Sehon notes that the teleological theory gives a mixed verdict in this case: Harris was rational, and so responsible and blameworthy—but only to a degree. Since it turns out that Harris had had an appalling upbringing, which arguably diminishes responsibility and blameworthiness, the verdict fits the case.

The problem is that, according to the teleological theory, anyone with the wrong values would come out as only partially rational, and hence only partially morally responsible and blameworthy. Sehon attempts to deflect this implication by arguing that someone who had a morally sound education and had greater rational capacities but ended up with “a warped view of other human beings” had “more capacity to see the error of his ways; he was less impervious to good reason” (166). Therefore, he would be more rational and hence more morally responsible than, say, Harris, on Sehon’s view. But although I agree that this would be the right verdict, I cannot see how the teleological view would yield it. First, an agent who has a greater capacity to understand the moral reasons that apply to him, and is even familiar with them, but disregards them strikes me as more rather than less impervious to good reason than someone who has lesser capacity or experience of the values. If we partially exonerate the latter, it is because we imagine that, with a sound upbringing and higher rational capacity, he might have acted better. But the former had both and yet doesn’t act better. Second, as we saw above, the teleological view says that “an agent’s responsibility is at least mitigated somewhat if the behaviour was not rational” (145). The criterion for rationality, freedom, moral responsibility, and blameworthiness on the teleological view, then, is not

an agent's rational capacities but the nature of his overall behavior: how rational that is. And, according to the theory, behavior that manifests "subhuman values" is thereby irrational and so not fully free, blameworthy, and so on.

While some readers will remain unconvinced by Sehon's position, its forceful and distinctive arguments are undoubtedly a welcome contribution to many of the debates that have dominated the theory of action in the past fifty years.

MARIA ALVAREZ
King's College London

Shaw, William R. *Utilitarianism and the Ethics of War*.
New York: Routledge, 2016. Pp. 196. \$155.00 (cloth); \$44.95 (paper).

Nonconsequentialist approaches to military ethics have enjoyed supremacy as long as the tradition has existed. It is the conventional wisdom that utilitarianism was at best moot on the subject or at worst downright hostile to widely accepted moral constraints on declaring and fighting war. Utilitarians have chafed at this rejection, though a thoroughgoing defense of utilitarianism's ability to contribute to discussions of the morality of war has been lacking. This is surprising, given that war has been perhaps the most consequential enterprise in human history and, as Shaw points out in his second chapter, given the attention that the classical utilitarians—including Bentham, James and John Stuart Mill, and Sidgwick—paid to the conduct of war. Where utilitarianism has established a beachhead, it is only on the margins of these discussions, for example, when Walzer grudgingly accepts a threshold utilitarian calculus into his doctrine of supreme emergency. In this atmosphere, which is at turns dismissive and hostile, it is refreshing to see a contribution to the literature as lucid, thorough, and well argued as William Shaw's *Utilitarianism and the Ethics of War*.

Shaw's project is to argue that it has been a profound mistake to count utilitarians out of these discussions and that, in fact, utilitarianism provides the most satisfying unifying account of the rules of warfare, both *ad bellum* and *in bello*. The book is compelling, and its arguments are executed with an impressive economy of words, but any project of such ambition cannot hope to answer every question decisively in 166 pages. I will make a few brief remarks indicating where I expect readers to remain suspicious. I will suggest where I think there is work left to be done for utilitarians, now that Shaw has provided a promising framework.

Shaw's chief contribution to discussions of *jus ad bellum* is the Utilitarian War Principle, which he defends in his third chapter and returns to repeatedly:

Utilitarian War Principle (UWP):

It is morally right for a state to wage war if and only if no other course of action available to it has greater expected well-being; otherwise, waging war is wrong. (47)

The Utilitarian War Principle is implied by utilitarianism, though this implication flows in only that direction—we can accept this principle without being